STM JOURNALS

Journal of
Remote Sensing & GIS

https://journals.stmjournals.com/jorsg

ISSN: 2230-7990 (Online)
ISSN: 2321-421X (Print)
Volume 15, Issue 3, 2024
September–December
DOI (Journal): 10.37591/JoRSG

Review JoRSG

Safe Travel: Road Accident Analysis, Severity Prediction, and Safe Route Mapping

Siddhesh Shivekar[1,*], Parth Surve[1], Prathamesh Walawalkar[1],
Asad Mujawar[1], Jignesh Patil[2]

Abstract

Road accidents pose a significant threat to public health, resulting in millions of injuries and fatalities annually. With an estimated 1.2 million lives lost and 20 to 50 million people injured each year, the escalating trend of traffic accidents demands urgent attention. To address this issue, specialists utilize advanced algorithms such as random forests to analyze historical road crash data, aiming to predict accident hotspots. By identifying patterns and trends within this data, our study aims to uncover the root causes of accidents, enabling proactive prevention measures. Through a user-friendly web application, individuals can access real-time information about accident-prone areas within a 250-meter radius of their location. Leveraging data analysis and visualization tools like Power BI, Excel, and SQL, we pinpoint deficiencies in the current road infrastructure and collaborate with authorities to formulate targeted safety strategies and regulations. Our goal is to enhance road safety, reduce accidents, and save lives.

Keywords: Big data, structured query language (SQL), Power BI, road accident, leaflet js, HTML, cascading style sheet (CSS), severity prediction, random forest, prediction, road map

INTRODUCTION

The increasing number of car purchases over the years has undeniably led to a concerning rise in the rate of accidents, resulting in significant losses, both in terms of lives and property. This alarming trend has now become a critical issue that demands immediate attention and recognition by traffic-related law enforcement organizations. To address this issue effectively, a comprehensive analysis has been undertaken, focusing on identifying the root causes and various factors influencing these accidents. Encompassing datasets containing over a million data points from diverse regions, this analysis has provided valuable insights into the dynamics of road accidents.

***Author for Correspondence**
Siddhesh Shivekar
E-mail: siddheshshivekar@gmail.com

[1]Undergraduate, Student, Department of Computer Engineering, Rajiv Gandhi Institute of Technology Mumbai University, Mumbai, Maharashtra, India
[2]Assistant Professor, Department of Computer Engineering, Rajiv Gandhi Institute of Technology Mumbai University, Mumbai, India

Received Date: July 05, 2024
Accepted Date: August 17, 2024
Published Date: September 10, 2024

Citation: Siddhesh Shivekar, Parth Surve, Prathamesh Walawalkar, Asad Mujawar, Jignesh Patil. Safe Travel: Road Accident Analysis, Severity Prediction, and Safe Route Mapping. Journal of Remote Sensing & GIS. 2024; 15(3): 39–44p.

These insights are invaluable for traffic-related law enforcement organizations, as they enable the development of targeted measures to mitigate the risks associated with accidents. enforcement agencies can implement evidence-based strategies and interventions aimed at reducing the occurrence of accidents and saving lives. This project serves as a crucial step toward enhancing road safety, as it equips authorities with the necessary information to make informed decisions and take proactive actions, ultimately leading to a safer and more secure road environment for everyone. It is imperative that these findings are embraced and translated into concrete policies and actions to address the pressing issue of road accidents effectively.

RELATED WORK

In the article, “Road Accident Analysis and Hotspot Prediction Using Clustering.” Patil et al. [1] utilized K-means, predicting accidents and offering safety advice via intelligent model. K-means clustering may not work well for road accident analysis with big data due to sensitivity to initial choices and difficulty handling complex accident patterns [1].

Road accident prediction and model interpretation could be done by using random forest and hybrid K-means. Yassin and Pooja [2] followed variable-based classification on road accident severity to create new features and classification for prediction, where external factors were not taken into consideration.

Proactive B for traffic accidents prediction is a proactive system to predict road accidents considering external factors, utilizing predictive modelling for infrastructure changes. Islam et al. [3] proposed system may require significant computational resources and expertise to implement effectively.

Discoveries highlight that mishaps or mischances are generally caused by individuals [4]. In expansion the condition of vehicles and foundation can influence the mischances and lead to passings. Tending to these circumstances ought to be of extraordinary significance for all nations and endeavors ought to be made to fortify security measures. The researchers advocate the creation and usage of laws and directions pointed at decreasing mortality rates.

Mohamed and Hassan [5] utilized a database from a fatality analysis reporting system to analyze the impacts of increasing registrations of light truck vehicles (LTVs) on fatal angular collision trends in the USA. They also investigated the total annual fatalities resulting from angular collisions and collision configurations. Time series modeling results suggested that fatalities in angular collisions are expected to increase over the next decade, potentially influenced by the projected overall growth of LTV proportions in traffic [5].

Wahab and Jiang [4] performed a crash investigation on the Ghana dataset utilizing Multilayer Perceptron (MLP), Portion, and Simple CART to distinguish exceptions and decide the centrality of bike crashes [4]. The authors utilize the Weka apparatus to compare and analyze the dataset and utilize Info Gain Attribute Eval to discover the factors that have the most prominent effect on cruiser crashes in Ghana. Hence, the simple CART show appears superior exactness than other classification models.

The calculation of the road safety composite index for the purpose of identifying and rating hotspots is a continuous research endeavor that is aided by Coll et al. [6]. The present approaches aim to minimize the shortcomings, and this work introduces an aggregation method for creating the composite road safety performance index. To prioritize initiatives for improving road safety, this innovative approach can also be seen as an intelligent decision support system for road safety performance evaluation [6].

By including driver-specific characteristics, Arbabzadeh and Jafari's [7] innovative data-driven approach to predicting traffic safety risk may be tailored for each individual driver. To create the predictive models, they used data from the Second Strategic Highway Research Program (SHRP 2) naturalistic driving study and elastic net regularized multinomial logistic regression. To improve the prediction performance regarding model predictors, they looked at the variables in the data set and carried out data preparation and feature engineering procedures. The model yields good results, and the concluding section discusses model expansions and adaptations for even better outcomes. Depending on how many warnings the model can provide given the driving conditions, two variants of the model are shown [7].

Joshua and Garber [8] revealed mathematical equations linking the annual number of truck-related accidents at a highway stretch with traffic and geometry characteristics. These links were obtained using repeated linear and Poisson regression analyses. According to these models, the number of truck-related

ISSN: 2230-7990 (Online), ISSN: 2321-421X (Print)

accidents at a particular highway stretch is influenced by the slope change rate (the absolute curve of slope changes in the vertical direction divided by the highway segment), average daily traffic, the percentage of trucks, and the speed difference between trucks and non-trucks [8].

At the segment level, Aguero-Valverde and Jovanis [9] investigated the impact of spatial correlation in models of the frequency of traffic accidents. To determine which segment neighboring structure is best for modelling crash frequency in road networks, several architectures are tested. For the spatial correlation terms, a full Bayes hierarchical approach is employed with conditional autoregressive effects. The significance of incorporating spatial correlation in road crash models is demonstrated by an analysis of crash, traffic, and roadway inventory data from a rural Pennsylvanian county [9].

PROPOSED SYSTEM

Our model integrates diverse factors including age, gender, atmospheric conditions, vehicle and road conditions, and the mental state of drivers to achieve optimal accuracy. Through the utilization of the random forest algorithm across comprehensive datasets, we strive to deliver precise predictions regarding accident causation across different regions in India. This empowers us to effectively pinpoint the factors contributing to accidents, facilitating prompt intervention to mitigate future incidents.

Algorithm

The process starts with random forest algorithm, an ensemble learning technique. This is used for classification and regression tasks.

- *Training Set:* The data is divided into a training set, used to teach the model, and a testing set, which remains untouched for later evaluation. For accident hotspot prediction using random forest, we have considered road conditions, intersection type, and vehicle type.
- *Road Condition:* Dry, wet, icy, or snow-covered roads affect accident likelihood due to varying traction and visibility.
- *Intersection Type:* Signalized intersections regulate traffic flow, while unsignalized ones rely on right-of-way rules, potentially increasing accidents. Roundabouts promote slower speeds.
- *Vehicle Type:* Differences in size, speed, and maneuverability among passenger cars, trucks, motorcycles, and bicycles influence accident patterns. The accuracy which we obtained on training set was 81% and, on our testing set, we were able to get an overall accuracy of about 70%.

Classifier Applied

By integrating these factors, random forest identifies hotspot locations where specific combinations of road conditions, intersection types, and vehicle types lead to accidents. This aids stakeholders in implementing targeted interventions to enhance road safety.

- *Testing Set:* The testing set, a portion of the dataset separate from the training set, is employed to evaluate the model's performance.
- *Classifier Applied:* A decision tree model is applied to both the training and testing sets to make predictions.
- *Training Model:* The model is trained using the training set, adjusting its internal parameters to make accurate predictions.

Model Comparison

- *Model Comparison:* Once trained, the model's performance is compared on both the training and testing sets to assess its generalization capabilities. This step helps identify potential overfitting or underfitting issues.
- *Model Interpretation:* The final step involves interpreting the model's behavior and results. This could include understanding which features are most influential or drawing insights from the model's predictions.

Web App

The application uses geofencing technology to highlight accident-prone zones on a map as red circles. It utilizes the Geolocation API (application programming interface) to track the user's location in real-time. When the user enters a predefined geofence area, the application triggers visual alerts via SweetAlert2 and audible warnings using the Speech Synthesis API provided by Google Text-to-Speech Service. Users can also set a destination, and the Leaflet Routing Machine library calculates and displays the shortest route. The application continuously analyzes spatial data, computing the distance between the user's location and geofence centers to detect if they are inside or outside risk areas. We have made use of spatial analysis. As the user's location changes, the app dynamically updates this distance computation.

The overall system flow is shown in Figure 1.

Data Gathering

Data was collected from Kaggle. This dataset provides a clear knowledge about the circumstances. The dataset has the following attributes:

- Coordinates
- Accident-prone areas
- Velocity
- Date and time
- Duration
- Location
- Atmospheric condition
- Condition of vehicle
- Potholes
- Traffic conditions
- Drunk drivers
- Vehicle type

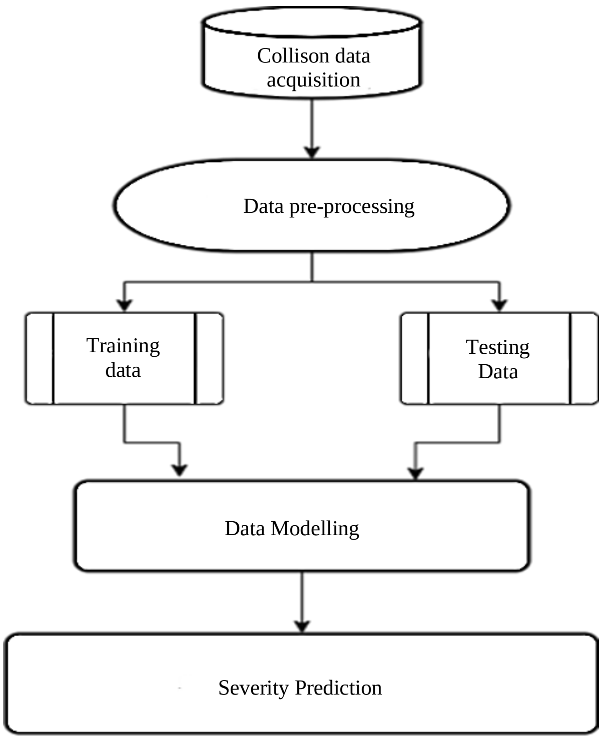

Figure 1. System flow of the safe travel system.

Data Modeling

Data modeling steps for severity prediction:

1. *Data Preprocessing:* Various mappings are applied to categorical variables, transforming them into more interpretable forms. For example, accident severity levels are mapped from numerical values to descriptive labels such as 'Fatal', 'Grevious Injury', etc. [10] The target variable (y) is assigned as the 'Accident_Severity_C' column, and predictor variables (X) are selected by dropping irrelevant columns.
2. *Train-Test Split:* The dataset is split into training and testing sets using the train_test_split function, with 20% of the data reserved for testing (test_size = 0.2).
3. *Model Training with Hyperparameter Tuning:* A random forest classifier model is instantiated with a random_state parameter set to 0. A grid search is performed to find the best combination of hyperparameters (n_estimators and max_depth) using 5-fold cross-validation. The best model is selected based on the highest mean cross-validated score obtained during the grid search.
4. *Model Evaluation:* The best parameters and corresponding score are printed. The best estimator from the grid search is fitted to the training data. Predictions are made on the testing data using the best model.

DISCUSSION

About 41% of the crucial factors in situations where the researchers assigned the cause to the driver were recognition failures (inattention, internal and external distractions, insufficient surveillance, etc.). Furthermore, judgement errors (driving aggressively, going too fast, etc.) accounted for around 34% of the significant reasons that the driver was blamed for, while performance faults (overcompensation, inappropriate directional control, etc.) accounted for 10%. The investigators also evaluated other crash-related elements, such as interior non-driving activity. As a matter of fact, roughly 18% of the drivers were involved in at least one activity that involved not driving inside [11].

The field of traffic safety has made significant progress in improving cars' crashworthiness, or their capacity to keep their occupants safe in the event of an accident. More primary prevention work is required in order to significantly lower the high rate of traffic-related fatalities and injuries (i.e., identifying strategies to prevent crashes by studying the events leading up to a crash.) The automotive sector has already invested a large amount of money in the study and creation of crash avoidance technologies for automobiles. In the fleet of newer model cars, several of the new features (traction control, lane-departure warning systems, Electronic stability control (ESC), etc.) are beginning to appear.

CONCLUSION

In conclusion, the analysis of road accidents using Power BI has provided invaluable insights into the various factors contributing to road safety and has paved the way for informed decision-making and targeted interventions. Through the power of data visualization, we have been able to explore the patterns, trends, and critical aspects of road accidents in a comprehensive and accessible manner.

The random forest model exhibited a significant discrepancy between training accuracy (81%) and test accuracy (70.47%), indicating potential bias or class imbalance within the training dataset. Analysis revealed no statistically significant correlation between the day of the week and the incidence or severity of traffic accidents.

As we move forward, it is essential to continue refining and updating our road safety strategies based on the ongoing analysis and insights. By doing so, we can strive to create safer road environments, prevent accidents, and ultimately, save lives.

Road accident analysis using Power BI serves as a testament to the power of data in addressing complex societal challenges, and it reinforces the importance of data-driven decision-making in our efforts to make our roads safer for everyone.

Acknowledgements

We deeply appreciate Dr. Sanjay U. Bokade, Principal, and Prof. S.P. Khachane, Head, Department of Computer Engineering at MCT's Rajiv Gandhi Institute of Technology, for providing us with the valuable opportunity to pursue our project, "Analysis and Prediction of Road Accidents."

Their consistent backing and motivation have been crucial in the successful execution of our project. We are immensely grateful to our esteemed project mentor, Prof. Jignesh Patil, whose guidance and expertise have been indispensable throughout every stage of this endeavor. His mentorship has not only honed our abilities but also inspired us to strive for greater accomplishments.

Additionally, we would like to thank our committed peers and supportive friends for their collaboration and assistance, which have been integral in bringing this project to fruition. Their contributions have enriched our journey and elevated the caliber of our work.

REFERENCES

1. Patil J, Patil V, Walavalkar D, Lobo VB. Road accident analysis and hotspot prediction using clustering. In: 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, July 8–10, 2021. pp. 763–768. doi: 10.1109/ ICCES51350.2021.9489074.
2. Yassin SS. Pooja. Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach. SN Appl Sci. 2020; 2: 1576. doi: 10.1007/s42452-020-3125-1.
3. Islam M, Patel D, Hasan AS, Jalayer M. An exploratory analysis of two-vehicle crashes for distracted driving with a mixed approach: machine learning algorithm with unobserved heterogeneity. J Transport Safety Security. 2023; 16 (7): 709–745. doi: 10.1080/19439962.2023.2248035.
4. Wahab L, Jiang H. A multinomial logit analysis of factors associated with severity of motorcycle crashes in Ghana. Traffic Inj Prev. 2019; 20 (5): 521–527. doi: 10.1080/15389588.2019.1616699.
5. Abdel-Aty M, Abdelwahab H. Modeling rear-ends collisions including the role of driver's visibility and light truck vehicles using a nested logit structure. Accident Anal Prev. 2004; 36 (3): 447–456.
6. Coll B, Moutari S, Marshall AH. Hotspots identification and ranking for road safety improvement: an alternative approach. Accident Anal Prev. 2013; 59: 604–617.
7. Arbabzadeh N, Jafari M. A data-driven approach for driving safety risk prediction using driver behavior and roadway information data. IEEE Trans Intell Transport Syst. 2018; 19 (2): 446–460. doi: 10.1109/TITS.2017.2700869.
8. Joshua SC, Garber NJ. Estimating truck accident rate and involvements using linear and Poisson regression models. Transport Plann Technol. 1990; 15 (1): 41–58.
9. Aguero-Valverde J, Jovanis PP. Analysis of road crash frequency with spatial models. Transport Res Rec J Transport Res Board. 2008; 2061: 55–63.
10. Ansodariya D. Kaggle. UK Road Safety (Accident) Analysis. [Online]. 2022. Available at https://www.kaggle.com/code/devansodariya/uk-road-safety-accident-analysis/input
11. National Motor Vehicle Crash Causation Survey: Report to Congress. National Highway Traffic Safety Administration. [Online]. 2008. Available at https://crashstats.nhtsa.dot.gov/ Api/Public/ViewPublication/811059

